



Use of Topic Knowledge in Spoken Dialogue Information Retrieval System for Academic Documents

Shinya Kiriyama †, Keikichi Hirose ‡, Nobuaki Minematsu †

† Department of Information Engineering, School of Engineering

‡ Department of Frontier Informatics, School of Frontier Sciences

University of Tokyo, Japan

{kiriyama, hirose, mine}@gavo.t.u-tokyo.ac.jp

Abstract

An efficient search function based on topic estimation was integrated to our spoken dialogue system for academic document information retrieval. The following two points were mainly studied: 1) to properly categorize documents (to be retrieved) into related topics, and 2) to facilitate retrieval process using topic knowledge. For the first point, a method was developed to calculate recursively the relevance scores of retrieval words and documents for topics. Effects of the recursive process were proved through experimental results; better classification of retrieval words and documents into topics was realized. As for the second point, retrieval range was limited into topics estimated from retrieval words. It was shown through experiments of retrieval task solving that necessary number of dialogue turns (therefore, period of dialogue) could be largely reduced by the range limitation; a smooth retrieval process was proved to be realized using topic knowledge.

1. Introduction

In these years, topic detection/tracking has been studied intensively aiming at such as automatic indexing of news speech [1][2][3]. While, we developed a spoken dialogue system for information retrieval on academic documents[4]. In the current paper, we introduced topic detection process into our system to realize an efficient retrieval. Topic (category) of documents which users are searching for is estimated from accumulated relevance scores between the topic and retrieval words included in user's input utterances. Here, 'retrieval words' are defined as words utilized for matching during the retrieval process. Also, 'retrieval field' henceforth denotes topic of documents retrieved. Efficient retrieving is realized by limiting the search area into the estimated retrieval field. Currently, four fields are considered: 'speech (processing),' 'image (processing),' 'communication,' and 'others (not included in the three fields).' The following two points were mainly taken into account:

1. Properly categorize documents into the most suitable retrieval field.
2. Realize an efficient retrieving by limiting the following dialogue into the estimated retrieval field.

For the first point, two types of relevance scores were used: one between a field and a retrieval word (*RSW*), and the other between a field and a document (*RSD*). We developed a recursive method to calculate these two types simultaneously. Using the method, for a given topic, the extraction of its retrieval words from the database and the classification of documents in

the database into the retrieval fields are carried out automatically. All documents are belonged into appropriate retrieval field based on the *RSD*.

As for the second point, the topic information obtained by the developed method was introduced into the dialogue management strategy and the visual output strategy of the system. These strategies were evaluated from the viewpoint of retrieval efficiency; to what extent dialogue turns and periods were reduced as compared to the case without control. In order to check overall usability of the system, subjective evaluation results were also examined.

2. Topic knowledge acquisition

Our original method for *RSW* and *RSD* calculation consisted of the following steps: 1) select retrieval words useful for the topic identification based on χ^2 -values (henceforth, 'distinctive words'), 2) for each selected word, calculate *RSWs* to the retrieval fields, and 3) compute *RSDs* for each document by accumulating *RSWs* of distinctive words appeared in the document. We newly introduced the recursive process to improve the performance. Several alternatives are checked for the selection of retrieval words and the treatment of compound nouns.

2.1. Calculation of relevance scores

Currently, the document database in the system consists of 1,859 articles, all of which include titles, keywords, and abstracts. Retrieval score calculation is done as follows:

1. Initial grouping of documents into 4 fields. Frequency (number of appearance) of word 'speech' in a document is counted and is normalized by its total frequency in the entire database. The normalized frequency was also calculated in the same way for words 'image' and 'communication.' The document is categorized to a field with the same name as the word with the largest normalized frequency: for instance, 'speech' field if word 'speech' takes the largest frequency. If non of three words appear in the document, it is categorized to 'others.'
2. Counting frequency x_{ij} of retrieval word w_i in documents belonging to field t_j . The counting is done for titles, keywords and abstracts of documents.
3. Calculating χ^2 -value defined by eq. (1) as an index on how retrieval word w_i relates to retrieval field t_j .

$$\chi_{ij}^2 = \frac{(x_{ij} - m_{ij}) \cdot |x_{ij} - m_{ij}|}{m_{ij}} \quad (1)$$

where m_{ij} is a-priori frequency estimated by eq. (2):



$$m_{ij} = \frac{\sum_{l=1}^L x_{il}}{\sum_{k=1}^K \sum_{l=1}^L x_{kl}} \times \sum_{k=1}^K x_{kj} \quad (2)$$

where K and L denote total number of retrieval words and field numbers (4, for the current system), respectively. As it is clear from eq. (1), χ^2 -value may take a negative value [3].

4. Selecting distinctive words. First, highest χ^2 -value is selected for each retrieval word. Then, retrieval words are sorted in the order with higher 'highest χ^2 -value,' and top N words are selected. RSW_{ij} of each extracted word w_i for field t_j is calculated by eq. (3).

$$RSW_{ij} = \frac{x_{ij}}{\sum_{l=1}^L x_{il}} \quad (3)$$

For retrieval words not selected by the above process, if their χ^2 -values take negative values for fields other than 'others,' they are assumed to be related to 'others' and are added to distinctive words. Otherwise, they are assumed to have relationship with none of 4 fields, and their χ^2 -values are set to 0 (excluded from the topic estimation process).

5. Calculating RSD s for each document. RSD_{dj} for document d and field t_j is computed by

$$RSD_{dj} = \frac{\sum_{l_d=1}^{N_d} RSW_{l_d j}}{N_d} \quad (4)$$

where w_{l_d} is l_d th word in document d , and N_d is the number of retrieval words which appear in document d .

6. Categorizing documents. Each document is categorized in the retrieval field with largest RSD .
7. Going to recursive process or not. If there is no change in the result of document categorization, finish the process. If not, going back to step 2.

2.2. Experiments

Experiments on document categorization were conducted by changing threshold N for distinctive word number (step 4, in section 2.1). For evaluating the procedures above, topic categorization rate (TCR) was employed, which was defined as rate of correctly categorized documents out of total documents. The correct categorization was obtained manually before the experiments. Three alternatives were investigated for the retrieval word selection, and inclusion/exclusion of compound words was checked for the calculation of RSD :

Selection of retrieval words: Based on the result of morphological analysis for titles, keywords, and abstracts of all the documents in the database, all nouns are selected as candidates of retrieval words. Compound nouns are detected as noun sequences sand-witched by non-noun words, and are included in the candidates. The following three cases are tested: **KA**: To use all the candidates as retrieval words (31,967 words), **KB**: To exclude words appearing only once in one document from the candidates (13,129 words), **KC**: To exclude words appearing only in one document from the candidates (7,975 words).

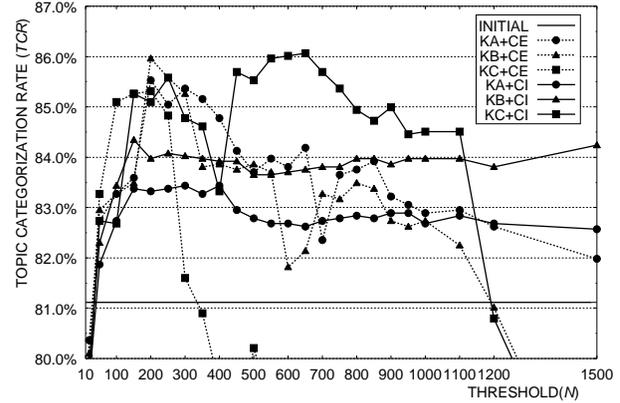


Figure 1: Topic categorization rates versus threshold N for various cases.

Treatment of compound nouns: When calculating RSD according to eq. (5), two cases are compared for compound words: to exclude them from calculation (**CE**) and to include them in calculation (**CI**). For both cases, RSW s of constituent nouns are included.

Results are shown in figure 1. The horizontal line at $TCR=81.1\%$ indicates initial TCR after step 1. Solid lines and dotted lines indicate results for case CE and CI, respectively. When the method is applied for other document database, the favorable case will be that taking high TCR in a wider range of N . From this viewpoint, the best set will be KC and CI combination, where TCR takes the peak value 86.1%, 5 point increase from the initial value, at $N=650$. The evaluation experiments in section 4 were done under this condition. For N ranging from 450 to 700, TCR exceeds 85.5 (%), which was more than 2 point increase from the baseline value obtained by our original method without recursive process.

2.3. Topic estimation

During the retrieval process, a user's interest may be placed on a retrieval field (or fields) and may change as the dialogue proceeds. The topic (field with user's interest) at time t was estimated using $C_j(t)$ defined as follows:

$$C_j(t) = \frac{\sum_{k=1}^{N_t} RSW_{kj}}{\sum_{l=1}^L \sum_{k=1}^{N_t} RSW_{kl}} \quad (5)$$

where N_t is the number of retrieval words included in user's inputs until time t , and L is the number of retrieval fields. The retrieval field(s) t_j with larger $C_j(t)$ than $1/L$ ($1/4$ for the current case) was presumed as the topic(s) at time t .

3. System development

3.1. Use of topic knowledge

The strategies of dialogue management and visual output were modified from those of original system to utilize topic knowledge obtainable through the process explained in section 2.

3.1.1. Dialogue management strategy

Dialogue management method of the original system was modified differently depending on the following conditions:



With estimated topic(s): When the retrieval field is not fixed during the preceding dialogue, two types of retrieval results, one without and the other with limitation to estimated topic, are offered to the user. Then, the system asks the user whether the following retrieval process should be restricted to the estimated topic or not. For instance, when the estimated topics are ‘speech’ and ‘image,’ the system makes the following reply; “As the result of retrieval, 20 documents are matched. 12 documents belong to ‘speech’ and 8 documents belong to ‘image.’ In which field you want to proceed the retrieving?” When the retrieval field is set already, retrieval results limited to the field are only shown. For instance, the system replies “5 documents are matched in the ‘communication’ field.”

With no estimated topic: When the retrieval field is not fixed, results for all fields are shown. While, when it is fixed, the system reply is the same as indicated above.

With direct topic specification by the user: The following retrieval process is done only in the specified field.

3.1.2. Visual output strategy

Two types of lists were possible as visual output: one with retrieved document titles and the other with retrieval words related to the retrieved documents (henceforth, ‘relevant words’). The system chose one of these lists for display depending on the number of retrieved documents. When the number was not more than 10, the (entire) title list was shown. On the other hand, when the number exceeded 10, the relevant word list was shown without the title list. In the latter case, the system asked the user to select relevant word(s) so that the number of documents was reduced to 10 or less.

Document title list: When only one topic (field) is estimated, retrieved documents belonging to the field are sorted depending on their *RSD* values. (Document with higher values comes earlier.) When two or more topics are estimated, the documents are first separated to their suitable field. Then for each field, the documents are sorted in the same way as above. Each document in the list is attached with a colored mark, which represents to which field the document belongs.

Relevant word list: The relevant words are sorted so that the word belonging to larger number of documents comes earlier. When plural words belong to the same number of documents, they are further sorted depending on their *RSW* values. When the retrieval fields are restricted, the relevant words not belonging to the fields are excluded from the list. Here, a relevant word is assumed to belong only to the field with the highest *RSW*.

3.2. System configuration

As shown in figure 2, the system consists of 5 modules as follows:

Speech recognizer: Continuous speech recognition parser “Julian,” which is the context free grammar (CFG) version for Japanese LVCSR (large vocabulary continuous speech recognition) engine “Julius” distributed as part of “Japanese Dictation Toolkit -1997 version- [5],” was adopted. This parser used a triphone model set included in the toolkit as acoustic models. Dictionary size was about 500 words. A task dependent CFG was constructed manually and utilized.

Dialogue manager: Recognition results are analyzed and system behavior is decided. In order to keep the topic knowledge, additional data structures are added to the information stack of the original system.

Data retriever: Documents in the database are selected through the matching process using retrieval words in user’s in-

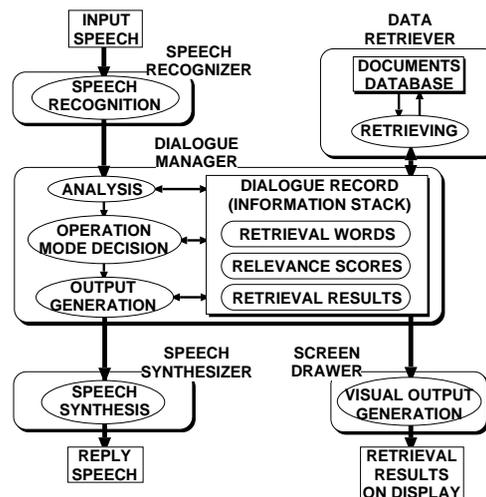


Figure 2: Configuration of the system.

put. When making the lists of titles and relevant words, they are sorted as explained in section 3.1.2.

Screen drawer: The following items are shown on display: recognition result of input speech, number of retrieved documents, estimated retrieval field(s), list of document titles, list of relevant words, and so on. The items on display changes as the dialogue proceeds.

Speech synthesizer: Two types of speech synthesizers are available: formant synthesizer formerly developed for our TTS system [6], and waveform concatenative synthesizer newly developed. The system evaluation experiments were conducted using the waveform concatenative synthesizer.

4. Evaluation

Eight Japanese subjects (five with experiences of using the original system beforehand and three without) were asked to evaluate two versions of the system: one using topic knowledge (the current version) and the other not using it (the original version). Each subject was asked to find out a document per a field (totally 4 documents) using one version. Also he was asked to find other 4 documents (one for each of 4 fields) using the other version. The selection of the version (which of current and original versions comes first) was randomized for documents and for subjects to cope with the ordering effect. A set of sentence examples acceptable by the recognizer and a set of available retrieval words were noticed to the subjects. Then a printed abstract was supplied, and the subjects were asked to find out the document from database which had the same abstract. After all the tasks were finished, they compared the two versions from the viewpoints listed in Table 1. A 5-rank scoring was adopted: point 2 if the current version is clearly better, point -2 if the original version is clearly better. Also they wrote down their impressions using the system. Table 1 shows the points of comparison in the 5-rank scoring.

4.1. Speech recognition performance

Word correct rate (*WCR*) and sentence understanding rate (*SUR*) were used as the measures for speech recognizer evaluation. *SUR* was defined as the ratio of the number of correctly understood sentences to the total number of input sentences during a dialogue. The overall values for *WCR* and *SUR* through the experiments were 91.4 % and 92.1 %, respectively. No com-



Table 1: Points of comparison.

Document title list:	
T1	Presentation of the list
T2	Sorting of the list items
T3	Contents of the list
Relevant word list:	
W1	Presentation of the list
W2	Sorting of the list items
W3	Contents of the list
Total impression:	
U1	Dialogue strategy
U2	Usability
U3	Preference

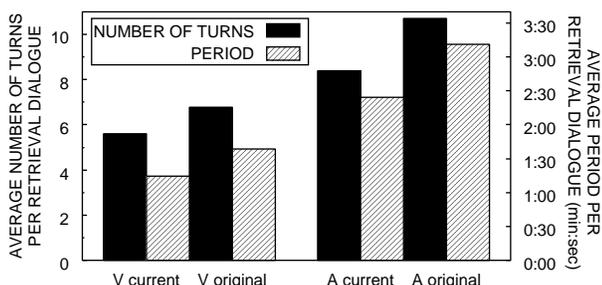


Figure 3: Average numbers of turns and average periods for a retrieval dialogue.

plaints were given from the subjects on the recognizer performance, indicating the evaluation results were not influenced by the mis-recognition.

4.2. Number of dialogue turns and period

We defined a ‘retrieval dialogue’ as a dialogue set used to finish a retrieval task (to find out a document). Average of the number of turns and the period per retrieval dialogue served as indices for retrieval efficiency. The number and period were counted for two cases: one for all the dialogue utterances (case ‘A’), the other only for valid utterances (case ‘V’). Here, ‘valid’ indicates that utterances to correct recognition errors and fixed-style utterances (such as “What kind of documents are you looking for?”) are not counted.

Figure 3 shows the results for ‘A’ and ‘V’ cases. In the figure, ‘current’ and ‘original’ mean the results for the current version and the original version, respectively. The results show that the average numbers of the current version are reduced from those of the original version by 17.5 % (for case ‘A’) and 22.0 % (for case ‘V’). The reduction was around 24 % when viewed from dialogue periods.

4.3. Results of subjective evaluation

Figure 4 indicates the results of subjective evaluation in terms of the 5-rank scoring. The results are indicated totally and separately for experienced and non-experienced users. All the items except W2 obtained higher scores for the current version, again indicating the validity of using topic knowledge.

The following impressions were also obtained: As for document title listings, the current version was evaluated as “The different coloring depending on the fields was helpful to get the retrieval results.” There were no clear comments related to T2 and T3. As for relevant word listings, the coloring of the current version was also evaluated. The current version fur-

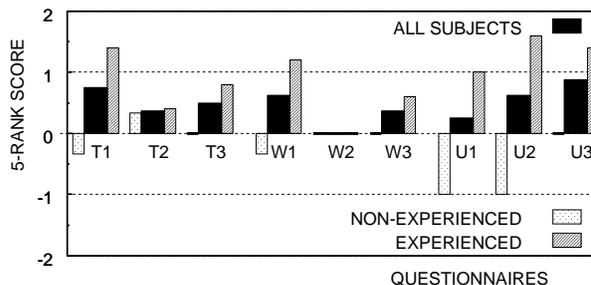


Figure 4: Results of 5-rank scoring.

ther evaluated as “It was easier to find necessary retrieval words from relevant words.” Several subjects did negative comments for the current system on point W2, indicating the necessity of further investigations. As for the total impression, the results U1 and U2 of the current version from non-experienced subjects were rather low. They commented “Simple function (to only use retrieval word list offered at the beginning of the experiment) is better.” This indicates that the retrieval strategy should be controlled depending on the user’s familiarity to the system. We have already examined both of system-oriented and user-oriented dialogue strategies, and the current system has a function of changing initiative between user and system dynamically. Therefore, the problem of user’s familiarity can be solved rather easily by starting the dialogue with the question to the user; such as, “Is this your first time to use the system?” The system can switch the initiative depending on the answer.

5. Conclusions

A scheme of facilitating the retrieval process was realized using topic knowledge in our spoken dialogue system on academic document information retrieval. A new method to calculate the relevance scores for retrieval words and documents for each retrieval field was proposed. Also a recursive method was developed to categorize each document to a retrieval field. The experimental results showed that 86.1% documents could be correctly categorized. Evaluation was conducted through the document search task. Both of objective and subjective evaluation results were favorable for the proposed scheme.

6. References

- [1] F. G. Walls et al, “Probabilistic Models For Topic Detection And Tracking,” Proc. ICASSP99, Vol. 1, pp.521-524, 1999.
- [2] P. van Mulbregt et al, “Text Segmentation and Topic Tracking on Broadcast News Via a Hidden Markov Model Approach,” Proc. ICSLP98, Vol.6, pp.2519-2522, 1998.
- [3] K. Ohtsuki et al, “Topic Extraction Multiple Topic-words in Broadcast-news Speech,” Proc. ICASSP98, Vol.1, pp.329-332, 1998.
- [4] S. Kiriya et al, “Development and evaluation of a spoken dialogue system for academic document retrieval with a focus on reply generation,” Proc. SST2000, pp. 32-37, 2000.
- [5] T. Kawahara et al, “Japanese Dictation Toolkit – 1997 version –,” J. Acoust. Soc. Jpn. (E), 20-3, pp.233-239, 1999.
- [6] K. Hirose and H. Fujisaki, “A System for the Synthesis of High-Quality Speech from Texts on General Weather Conditions,” IEICE trans. Fundamentals, Vol.1E76-A, No.11, pp.1971-1980, 1993.